6.033 Impact Framework, Spring 2022

Based on the Ethical Computing Protocol, created by Abby Everett Jaques and Milo Phillips-Brown

Preamble

The 6.033 Impact Framework is based on the Ethical Computing Protocol, created by Abby Everett Jaques and Milo Phillips-Brown. If you've seen that protocol in other classes, this framework is essentially a lighter-weight version of Steps 1 and 2, which we believe will be more conducive to using during your design project.

This framework might appear overwhelming at first! Don't worry about getting everything "right" on the first try, or on any try; this framework is simply a tool to help you think through your design decisions. For your first DP deliverable, we're only focusing on Steps 1.1-1.3. We'll move on to later steps in later deliverables, once you've gotten familiar with the earlier steps.

Introduction

The decisions we make when we design computer systems affect more than just the users and the developers, and moreover, they do not affect all users equally.

This framework contains a series of prompts and other tools to help you work through questions of impact. It is meant to be iterative and collaborative, and can be applied at lots of different points in the design process. For instance, one can assess the impact of the entire system, of a particular design decision, etc. As you work through it, it may bring up questions that you can't answer yet! That's okay; the important thing is to keep thinking about these questions.

You should use this framework as a tool for completing your Design Project deliverables (we'll typically tell you how to use the framework for each deliverable). You do **not** need to turn in any formal responses to the below prompts unless we ask for them. It's useful at multiple steps during the design process: at the start, when you make a big design decision, as you begin your evaluation, etc.

You'll also see this framework pop up, both formally and informally, in other areas of the class. For example, in recitation we often ask you who is impacted by a system, or what possible futures a system might lead to. That's all part of the Impact Framework.

Step 1: Envision Futures

The first step of the protocol is meant to help you understand the full range of possibilities of your design at a relatively high level.

1.1 What and Why?

Answer the following questions:

- **Describe what you will produce with this project, in 2–3 sentences of plain language.**
- **Explain why it's worth producing, in 2–3 sentences of plain language.**

To help you answer these questions, think about the following:

- Who is your project for? Who isn't it for? Why should each group care?
- What is the problem you're trying to solve? Are there existing solutions? If so, what's different about yours?

1.2 The Future

What's the best-case scenario for your project? What's the worst-case scenario?

You should have several answers to those questions. One best-case scenario might have to do with how your project is received by users; another might be about how it advances the state of the art. One worst-case scenario might have to do with enabling bad actors; another might be about your career.

To help you think through these scenarios, consider the following questions:

- Think about novel uses: What kinds of domain-hopping are possible? You have a certain kind of use in mind, for a particular area of life or kind of activity. But what else could your project be used for? What would be different if your project was used in ways you don't intend?
- Think about bad actors: How could someone appropriate your project for malicious purposes? What would happen if they did?
- Think about failure modes: Nothing works perfectly all the time. What can go wrong with your project, and how? What will be the effects when failures occur?
- Think about differences in benefits and harms to people. Could a member of a group affected by your project—especially a member of a group that's been historically disadvantaged—show you that there's something about your project that affects them differentially? How?

1.3 Who?

You've already started to think about how different groups are affected by your project; let's refine that. **What *domains* are impacted by your project?**

You can think about domains as a series of nested and/or overlapping circles with you at the center. You might start with you, your team, etc. These groups are contained in large structures: communities, cities, industries, economies, and so on.

You might include:
- Companies where your system might be used
- Industries where your system might be used
- Communities where those companies/industries have effects

At this stage, it might not be clear yet exactly *how* your project will affect these domains. That's okay! The goal is just to start thinking about who's affected.

1.4 How?

The different futures you thought about previously will affect different domains differently; the best-case scenario for one domain might not be the best-case scenario for another.

We can capture this idea with a *Change Matrix*: a table that describes how your project might go well or badly to any degree, for each of your domains of change.

Generate as many possibilities as you can think of, filling in the cells. It may be helpful to think about different timescales: some short-term downsides generate longer-term upsides, and vice versa. You'll come back to this Change Matrix in Step 2 of the framework.

> *Note from the 6.033 staff: We've found it useful to start with five rows: Best Case / Pretty Good / Neutral / Pretty Bad / Worst Case. It might be tempting to start with just three: Best / Neutral / Worst. But pretty quickly you'll find some gradations between Best and Neutral, and between Neutral and Worst.*
>
> *It's okay to use more rows if that feels helpful to you! The original Ethical Computing Protocol suggests seven instead of five.*

	Domain 1	Domain 2	…	
Best Case				
Pretty Good				
Neutral				
Pretty Bad				
Worst Case				

To help you think through some of these scenarios:

- What will your family brag about when they talk about your project? What will they avoid mentioning?
- What would make people fight to be put on the project? What would make them refuse?
- How could your project end up as a feel-good story on the nightly news? How could it end up as the subject of a Pulitzer-Prize-winning exposé?
- If a historian writing a hundred years from now talked about your project, what do you hope they'd say? What would make you hope they didn't mention your name?
- What would the plot of a Hallmark movie about your project be? What about a Black Mirror episode?

Step 2: Surface Values

At this stage, you've thought about the domains affected by your project, and the impacts of various outcomes on those domains. Step 2 is all about refining this process: being more rigorous in our understanding of what exactly is good and bad about the futures you've imagined, and for whom.

2.1 Who? Again

The futures you envisioned aren't just good or bad; they're good or bad for someone. It's time to be explicit about who those someones are.

Anyone who's affected by your project is a stakeholder. That includes everyone who appears in your Change Matrix, in every domain: from you and your team to your users, to competitors, regulators, and even people who may have no direct contact with your project, but who gain or lose something because of the difference your project makes to the world.

We'll start by refining our idea of who's impacted by your system.

- **Look back at your Change Matrix. For each possible future** (i.e., each cell in the matrix)**, list all the people and groups who appear in it.** In some cases, this may be an expansion of your domains — e.g., listing the individuals on your team instead of "DP Team" — or you may have uncovered impacted people who aren't represented in any of your domains.
- **When you list a group, ask yourself if there are subgroups that will be affected differently. If so, add them.** For example, users may have quite different experiences depending on factors like age, race, gender, socio-economic or health status, geographic location—or even something like height, or hand size, or familiarity with Star Trek. So ask yourself what it is that makes a future work well or badly for a group, and then think about whether subgroups will have different experiences.

As you work through this, don't get too caught up on whether something you classified as a domain is actually a stakeholder — that doesn't matter. The goal is to understand all of the people impacted by your project.

To help you think through possible stakeholders, here are some starting points:
- Workers at companies where your system is/isn't deployed
- People living in towns that have/don't have your system
- Companies that do/don't deploy your system (subgroups: small, medium, large)
- People in communities where those companies/industries have effects (think about subgroups, too)
- People who might be put at risk by your system, if any
- People who might experience secondary effects of your system (e.g., from some sort of disruption of operations

2.2 How? Again

Now that you have a detailed list of stakeholders, it's time to understand how they are affected. The end-goal of this step is to refine your Change Matrix, noting that you now have a more precise idea of the people/groups affected (thanks to Step 2.1); the Moral Lenses tool will help you do this.

Moral Lenses

To this point, we've been relying on a loose, intuitive sense of things going well or badly. But it's time to get clearer about what exactly we mean when we think about the impact of our system on stakeholders.

The Moral Lenses are three ways of looking at a possible future, each of which brings different aspects of that future into view. Using them is a way to understand in what sense a future is good or bad for someone.

Working through each lens in turn is also a way of doing a kind of moral due diligence —making sure we are taking all the morally significant aspects of a project into account.

Lens 1: Outcomes

- What's different for stakeholders after your system exists, as compared to the starting state?
- What good or bad things do they have more or less of?

Those costs and benefits can come in the form of just about anything: money or time or power or prestige; opportunities, abilities, or liberty; pleasure or pain; cupcakes or cat videos.

Many or even most of the futures you envisioned probably speak the language of outcomes. But this isn't the only way things can be good or bad.

Lens 2: Process

- Were stakeholders given due care, respect, and consideration in the process of producing the outcome?

This lens is motivated by the thought that sometimes the ends don't justify the means; there can be good outcomes that we shouldn't pursue, because what it would take to achieve them would be unacceptable.

If I notice spinach in a colleague's teeth, I shouldn't reach over and pick it out myself; if a crying baby is causing passengers on a long-haul flight to become increasingly agitated, I shouldn't relieve their suffering by tossing the baby out of the plane.

Lens 3: Systems

- How are benefits and burdens—both in Outcomes and in Process—distributed among stakeholders? Is the distribution fair?

This lens draws our attention to the way that it's not just whether there are net benefits, or whether the process was appropriate; it also matters what the patterns are. If all the benefits flow to one group, while all the burdens are borne by a different group, that can be morally significant, and it certainly requires justification.

So we want to know: are background conditions and historical patterns of marginalization, inequality, and oppression reduced, or reinforced?

Change Matrix Revisited

The possible futures in your Change Matrix are labeled according to degrees of general goodness or badness; now you will get more precise, saying for whom a future is good or bad, and in what way. Start with the key stakeholders flagged in your Stakeholder Inventory, and include others to the extent you can.

- For each cell in your Change Matrix, indicate which stakeholder(s) it's ideal for—and any stakeholder(s) it's not ideal for. (The cells of your change matrix might get large! Feel free to change the format into whatever works for you; you don't have to keep it as a table with cells.)
- Use the Moral Lenses to help you think about ways things can be good and bad for stakeholders that may not be explicit in your initial description of the possible future. Add that information to your cells.

Using the Impact Framework

At this point, you've spent a lot of time thinking about how the decisions that you've made in your design might impact different groups of people. You may have made changes in your design in response to this exercise — for example, rejecting a small performance gain that would also have negative impacts on a particular group. In your DP deliverables, you should use the conclusions you've drawn from this framework as part of your justification for your design decisions.